WHAT IS CLAIMED IS:

1. A system for compressing text using variable length codes, the system comprising:

a memory device configured to store a set of variable length codes for a plurality of languages for compression of text, wherein the set of variable length codes includes variable code lengths based on language features; and

an encoder coupled to the memory device, the encoder configured to receive text in at least one of the plurality of languages, to generate a compressed text by assigning a code to each word in the text based on codes from the set of variable length codes that are associated with the at least one language of the text and to generate at least one header to be inserted in the compressed text, the header including information regarding the location in the compressed text of a subsequent change in code length.

2. A system according to claim 1, wherein the language features include word lengths and frequency of occurrence of words.

3. A system according to claim 1, wherein the information included in the header includes a distance to a subsequent change in code length.

4. A system according to claim 1, wherein the information included in the header includes a distance to a subsequent header in the compressed text.

5. A system according to claim 4, wherein the distance is a maximum distance between headers.

6. A system according to claim 1, wherein the set of variable length codes is generated using Huffman encoding.

7. A system according to claim 3, wherein the distance is measured based on text delimiters in the text.

8. A system according to claim 1, wherein a header is associated with each code in the compressed text that is associated with a change in code length.

9. A system according to claim 1, wherein the encoder is further configured to identify a character string in the text that does not have a corresponding code in the

set of variable length codes and to tag the character string to indicate it is not compressed.

10. A system according to claim 1, wherein the text is received as a continuous stream of text.

11. A system according to claim 1, wherein the encoder is configured to apply a second compression process to the generated compressed text.

12. A system for compressing and decompressing text using variable length codes, the system comprising:

a first memory device configured to store a set of variable length codes for a plurality of languages for compression of text, wherein the set of variable length codes includes variable code lengths based on language features;

an encoder coupled to the first memory device , the encoder configured to receive text in at least one of the plurality of languages, to generate a compressed text by assigning a code to each word in the text based on codes from the set of variable length codes that are associated with the at least one language of the text and to generate at least one header to be inserted in the compressed text, the header including information regarding the location in the compressed text of a subsequent change in code length;

a second memory device configured to store the set of variable length codes for a plurality of languages for decompression of the text; and

a decoder in data communication with the encoder and coupled to the second memory device, the decoder configured to receive the compressed text, to generate a decompressed text by identifying a word associated with each code in the compressed text based on the set of variable length codes stored in the second memory device;

wherein the decoder identifies changes in code length based on the at least one header included in the compressed text.

13. A system according to claim 12, wherein the language features include word lengths and frequency of occurrence of words.

14. A system according to claim 12, wherein the information included in the header includes a distance to a subsequent change in code length.

15. A system according to claim 14, wherein the distance is measured based on text delimiters in the text.

16. A system according to claim 12, wherein the information included in the header includes a distance to a subsequent header in the compressed text.

17. A system according to claim 12, wherein a header is associated with each code in the compressed text that is associated with a change in code length.

18. A system according to claim 12, wherein the encoder is further configured to identify a character string in the text that does not have a corresponding code in the set of variable length codes and to tag the character string to indicate it is not compressed.

19. A system according to claim 18, wherein the decoder is further configured to provide the tagged character string as original text.

20. A system for decoding compressed text using variable length codes, the system comprising:

a memory device configured to store a set of variable length codes for a plurality of languages for decompression of the text, wherein the set of variable length codes includes variable code lengths based on language features; and

a decoder coupled to the memory device, the decoder configured to receive the compressed text having a plurality of codes and at least one header, to generate a decompressed text by identifying a word associated with each code in the compressed text based on the set of variable length codes stored in the second memory device;

wherein the at least one header includes information regarding the location in the compressed text of a subsequent change in code length;

wherein the decoder identifies changes in code length based on the at least one header included in the compressed text.

21. A system according to claim 20, wherein the language features include word lengths and frequency of occurrence of words.

22. A system according to claim 20, wherein the information included in the header includes a distance to a subsequent change in code length.

23. A system according to claim 20, wherein a header is associated with each code in the compressed text that is associated with a subsequent change in code length.

24. A method for compressing text using variable length codes, the method comprising:

receiving text to be compressed;

identifying a language of the text;

generating a compressed text by assigning a code to each word of the text using a set of variable length codes associated with the language of the text;

identifying each change in code length in the compressed text; and

inserting at least one header in the compressed text, the at least one header including information regarding the location in the compressed text of a subsequent change in code length.

25. A method according to claim 24, wherein a header is associated with each code in the compressed text that is associated with a subsequent change in code length.

26. A method according to claim 24, wherein the text is received as a continuous stream of text.